Digital Preservation: Planning, Strategies and Trends

Satish Kumar,
Information Scientist,
ARIES,
Nainital, UK

ABSTRACT

Digital preservation is currently a major challenge for the information management, technological and scientific communities in all domains. It is also crucial at the organisational level with special pressure for all institutions with a responsibility for preservation, such as libraries and archives. Digital preservation is a very important part of the creation and management of any digital collection. Without the proper preservation a digital collection can easily become inaccessible and therefore useless in just a matter of a few years. Terms such as medium preservation and technology preservation are widely used when discussing issues related to the preservation of electronic records.

This paper discusses long-term archiving and long-term access to digital documents, with an emphasis on strategy for preservation. Selecting materials for digital preservation depends on whether the materials are both valuable and endangered, whether appropriate digitization procedures and standards for these materials exist, and whether copyright allows reasonable access for educational and research purposes.

Keywords: *Digital information, , Digital archiving, LOCKSS, , E-resources*

Introduction

'Digital' and 'Preservation' are two terms that encompass wide areas of interest in information management. The digital "factor" changes the nature of content, its forms of representation, media and usage; having deep implications for preservation. Preservation has long been one of the core functions of libraries and archives. It can be examined at two major levels: issues of meaningful collection and maintenance of content and, in a more strict sense, aspects of stability of the media that support such content.

A digital preservation strategy is a well-measured and documented approach to the preservation of digital items. For collecting archivists, the purpose of such an approach is to make sure that access to the born-digital archives accessioned by a repository can be maintained indefinitely (Helen H., 2006). For repositories responsible for personal digital archives this will probably include material created on long-past, recent, current and future computers and devices. This is a major challenge for digital archivists because the technological landscape evolves so rapidly: new hardware, software and removable media are developed and adopted on a regular basis, and new versions often only have limited backwards compatibility. Obsolescence therefore poses a major threat to the survival of digital records and must be addressed in a preservation strategy.Digital archivists must also be able to demonstrate to user that a preserved digital object is authentic: that no accidental or intentional changes have occurred; and that the stated date of creation or receipt, identity of the author and process that created the object is all verifiable. Preservation strategies are all based on

the principle that the preserved digital object should be identical in all essential respects to the digital object which left the creator's computer or other device (Kirchhoff, A. J., 2008). This means that it is important to understand what is 'essential' in order to protect those aspects of a record which are essential and to measure the success of preservation interventions.

DIGITAL PRESERVATION

Preservation for posterity has always been one of the main responsibilities of libraries. Traditional preservation has been an extension of the custodial function, under which importance was given to preservation of the '*physical form*' – manuscripts, books, sculptures, gramophone records, etc. In the digital world, traditional preservation concept is undergoing a transformation, due to the special features of digital resources.

The **UK CEDAR** project has given the following definition of '**digital preservation'**: "*Storage, maintenance and access to a digital object over the long term, usually as a consequence of applying one or more preservation strategies.*" (Digital Preservation coalition, 2002).

Digital documents are created using computer systems, and depend on them for continued access. Hardware and software systems become outdated very rapidly, leading to technological obsolescence. Digital preservation, therefore, involves taking measures to ensure that the digital documents will remain available, usable and authentic in the future when the applications and systems, which were used to create and interpret them, might no longer are available.

Preservation of digital resources is much more complex than that of non-digital material such as paper. *For example*, when a book is preserved, all aspects of the book are preserved – its physical presence, format, layout and contents, since all these individual elements are inextricably linked. Digital formats on the other hand can be easily separated into individual elements; hence more efforts are needed to preserve them as a whole. *For example*, one can retain the content of the digital document while losing the layout (converting HTML to PDF format); or one can maintain its physical presence i.e. the computer file, but fail to preserve its readability. One can think of the following different levels at which digital resources can be preserved:

- ***Preserve the physical presence***: Also known as '***bit preservation'***, under which the computer file is preserved.
- ***Preserve the content***: Since actual information content is the most crucial component, one can provide access to it at its lowest level such as ***ASCII text***, without the add-ons such as font variations or layout features.
- ***Preserve the presentation***: The original format or layout may include different font faces and sizes, columns, margins, the use of white space etc. that carry out functions such as emphasizing important parts of the text, improving clarity, etc. In many formats of digital documents (e.g. SGML, XML), the layout specifications are separate from the content. To maintain the original look of the document, these layout specifications must also be preserved.
- ***Preserve the functionality***: Special efforts are required to preserve the functionality of digital documents, which might contain multimedia components, exist in hypertext format and have navigation functions such as toolbars, keyword search or interactive tables of contents.

Digital preservation is the series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value (McDonald J. and Van de Velde, E., 2004).

In order to preserve materials on a scale commensurate with mass storage capabilities and in formats that are accessible and usable, it is necessary to articulate some basic requirements.

- ***User's perspective***: User's expectation is always changing, yet users, especially research scholars need both traditional documents and electronic documents of old information and current information.
- ***Institutions's responsibility***: Libraries, archives and other custodians have responsibility for their any properties. So institutions should plan for digital materials including their maintenance, preservation and distribution.
- ***Mission of parent institution***: First object of libraries, archives and other custodians is to satisfy the user's expectation and user's requirements. They should *preserve all materials in all formats*.
- ***Storage media***: Storage media is having different formats such as text, data graphics, video and sound, different storage capacity like floppy disk, CD-ROM (R/R-W), DVD ROM (R/R-W), Pen drive, External Hard disk drive etc.(Beagrie, N. and Jones, M., 2002).

Selection for Preservation

"Selection" is another important issue in digital preservation. The huge quantity of information being produced digitally, its variable quality, and the resource constraints on those taking responsibility to preserve long-term access make selectivity unavoidable for archiving.

Traditionally, lack of selection for preservation may not necessarily mean that the item will be lost, but in the digital environment non-selection for preservation will almost surely mean loss of the item. Although not all resources need to be preserved forever, some will not need to be preserved at all, others will need to be preserved only for a defined period of time, and a relatively small sub-set will need to be preserved indefinitely. Selecting materials for digital preservation depends on three criteria:

- Whether the materials are both valuable and endangered;
- Whether appropriate digitization procedures and standards for these materials exist;
- Whether copyright allows reasonable access for educational and research purposes.

However, just as the web allows anyone to publish information on it without having to obtain any formal permission, it also allows its removal. The library must capture relevant free as well as paid web-based resources whenever they are available, in order to safeguard against their possible loss from the web, so that they can provide continued access to these resources to its users (Obaiah, B. and J., Francis, 2008). Though libraries will have to cater to remote access to resources (online journals, online full-text databases etc.) as a part of collection development, it would be desirable to download and preserve selected core resources on local servers.

Degree of preservation

A digital object, such as the kind of typical object found in a personal archive (e.g. a speech written in a word-processed document, a budget in a spreadsheet, or a digital image of a family) is defined in the PREMIS *Data Dictionary* as a discrete unit of information in digital form which is comprised of three different levels:

1. **Bit stream:** in its simplest form, a digital object consists of a bit stream, i.e. an ordered sequence of bits (binary zeros and ones). This binary inscription will usually be stored on a physical medium of some kind. A computer system with the correct combination of hardware or software

translates the bitstream into something meaningful.

2. **File:** a named and ordered sequence of bytes known by an operating system. The format of a file is laid out in the format specification, which transforms the file from its binary ones and zeros into something which makes sense to a user, i.e. stipulating the proper encoding, sequence, arrangement, size and internal relationships which enable the construction of a valid file of the relevant type (e.g. jpeg 2013 a file). This level represents the transformation of the input bit stream into output for presentation purposes; the physical medium on which the bit stream is inscribed is therefore of no consequence at this level.
3. **Representation:** denotes the set of files needed for a complete and reasonable rendition of an Intellectual Entity. This is defined by PREMIS as a coherent set of content that can reasonably be described as a unit; this is essentially a conceptual object, or something that a human can understand as a meaningful unit of information, e.g. a website, a report, a photograph. An Intellectual Entity or conceptual object may have one or more digital representations or encodings; *for instance*, the text of a politician's speech might be saved as both a Microsoft Word document and a PDF file. The underlying encoding of each will differ considerably, but the textual content of each item is identical.

Simply preserving the bit stream therefore does not guarantee ongoing access to a digital object. The digital object has an existence separate from the medium on which the bit stream is inscribed and successful preservation is only complete when all the significant properties are maintained and the digital object can be displayed in a meaningful and understandable form. In some cases, this might mean that only the intellectual content (e.g. the text of a word-processed document) is preserved, and that original formatting and layout is not retained in the preserved object. However, in other cases (notably, but not exclusively, complex objects such as interactive resources) it may be important that the 'look and feel' of the original object, and its functionality, is retained or recreated as part of the preservation process.

Integrating digital preservation in library

Long-term preservation is an important mission for a digital preservation system in a library. It ensures that the digital collection the library builds will be preserved for future users, regardless of the vagaries of institutional funding and decisions made by publishers (Rebecca, P., 2013). However, that is only half of the story. Unless users can access preserved content when it is no longer available from the publishers, it is like locking print media in a sealed vault with no key for safekeeping. The content might be safe, but it is certainly not useful.

Mission

The other mission for a library's digital preservation system is to make the preserved content available whenever it is needed. That may occur at any time for one of many reasons: there could be a temporary interruption of internet access, the publisher may have blocked access to the library because of a mistake or a dispute, or the library may simply have canceled their subscription and permanently lost access. Regardless of the reason, the digital preservation system must be ready immediately to serve its content in place of the publisher's.

Strategy/Techniques

A digital preservation strategy is a well-considered and documented approach to the preservation of digital objects. For collecting archivists, the purpose of such a strategy is to ensure that access to the born-digital archives accessioned by a repository can be maintained indefinitely. For repositories

responsible for personal digital archives this will probably include material created on long-past, recent, current and future computers and devices. This is a major challenge for digital archivists because the technological landscape evolves so rapidly: new hardware, software and removable media are developed and adopted on a regular basis, and new versions often only have limited backwards compatibility. Obsolescence therefore poses a major threat to the survival of digital records and must be addressed in a preservation strategy.

Different strategies are being used to carry out preservation of digital resources. Some of the significant ones are:

- **Migration:** Under this method, there is a periodic transfer of digital materials from one hardware/software configuration to another. *For example*, file formats are converted into compatible new formats as soon as the original formats face the risk of becoming obsolete (e.g., HTML 3.2 to HTML 4.02 or .doc to .docx)
- **Emulation:** It involves retaining information about how a digital object was created and accessed so that future access can be accurately and authentically reproduced. Emulation retains the functionality, look and feel of the original document.
- **Technology Preservation**: This involves preservation of the technical environment by conserving the software and hardware needed for interpreting digital information.
- **Conversion of data to standardized format:** Data could be transferred to *ASCII format*. It retains information, but not the structure or functionality.
- **Universal Virtual Computer (UVC):** One of the latest techniques to be explored is the UVC-based preservation method, which allows digital objects to be reconstructed in its original appearance anytime in the future, a programme or emulator is written to carry out this interpretation in the machine language of the UVC at the time the record was archived (without requiring any knowledge of the future target machine). The data can be stores in any format and the knowledge required to decode it is encapsulated in the UVC programme (format decoder). The UVC concept consists of the UVC itself, a logical data scheme with type description, the UVC programme and the logical data viewer.

File Formats Selection for Digital Preservation

File formats encode information into forms that can only be processed and rendered comprehensible by very specific combinations of hardware and software. The accessibility of that information is therefore highly vulnerable in today's rapidly evolving technological environment. This issue is not solely the concern of digital archivists, but of all those responsible for managing and sustaining access to electronic records over even relatively short timescales. According to Lavoie, Issues to consider when selecting file formats for long-term preservation include:

- Is it defined by an international, national or publicly available standard?
- Is the quality of the specification adequate?
- How widely has the format been adopted as a preservation format?
- Is it backwards compatible?
- Is it independent of any specific hardware or software environment?
- Does it have good metadata support (i.e. metadata providing technical and provenance information which is generated by the creating application, entered manually by the record creator, or a combination of these)?
- Does it have a good range of functionality without being too complex for the purpose?

- Is it easily convertible into other formats (for migration purposes)?
- How well does it retain the formatting and other significant properties of converted digital objects?
- How stable is the format?
- How proven is it in terms of longevity?
- Does it include an error-detection facility? (Lavoie B., 2004).

One other approach to selecting preservation formats is provided by OCLC's INFORM Methodology, which measures the preservation durability of digital formats. It compares formats and preservation approaches, and provides a risk management-based means of tracking what might be lost over time if particular preservation actions are taken; the digital archivist can then make decisions about preservation strategy based on this risk assessment (Reich, V. and Rosenthal, D., 2001). If a digital repository chooses to convert digital objects to one or more standard formats, there are a number of ways for consideration; examples for text-based documents include:

Extensible Markup Language (XML): this is not a format, rather a general-purpose markup language for describing the structure and meaning of data. It is an open standard defined by the World Wide Web Consortium and is independent of specific applications. Preserving digital objects which have been created using XML in accordance with a standard DTD or Schema is straightforward. Converting other digital objects to XML is one kind of migration approach but with its own limitations.

OASIS Open Document Format for Office Applications: this is an open, XML-based, format for office files, such as word-processed documents or spreadsheets. It has been adopted as an international standard (ISO/IEC 26300) and offers a suitable format for the preservation of digital documents created in proprietary office formats like those generated by Microsoft Office.

Portable Document Format Archive (PDF/A): this is a constrained version of Adobe's PDF version 1.4 which has been adopted as an international standard (ISO 19005-1). It is preservation-friendly. Its specification is openly available. It eliminates elements likely to complicate decoding and accelerate obsolescence (e.g. audio and video elements, or encryption, etc., which are sometimes used in other PDF formats); it is self-contained (i.e. can be displayed without any reliance on information from external sources); and support for embedding metadata is very good. Records saved to this format have a look and feel which is fundamentally one of text and images designed to fit a particular page size. However, it preserves static visual appearance only, so it is not suited in cases where functionality or logical structure needs to be preserved.

For images include:

Tagged Image File Format (TIFF): a format used for raster (i.e. pixel-based) images. It is widely adopted and supported by most image processing and viewing applications, and it supports sophisticated colour management features. Many repositories consider TIFF to be the best option for preserving images, and it is often used to store archival masters of digitised images.

Joint Photographic Experts Group (JPEG): a widely used format to represent continuous tone images (e.g. photographs and grey scale images). It is defined by an international standard (ISO 10918).

There are also widely accepted options for sound formats (e.g. WAVE LPCM or MP3_FF) and moving image formats (e.g. MPEG-2, MPEG-4_AVC). Many digital repositories publish details of the preservation formats they support, where more information about accepted formats can be found.

An important consideration when selecting a preservation format is how successfully the chosen format embodies the essential attributes, or significant properties, of the original digital object.

Comparison of Six major e-Journal Archiving Programmes

The following six programmes have been selected for comparative study. They are operated by organisations with excellent credentials in this field. At this early stage and given the varied factors which need to be taken into account to meet the needs of individual institutions, none completely deal with all requirements. It may be helpful to see the different features, pros and cons of these e-journal archiving programmes to choose a better one.

LOCKSS

LOCKSS (Lots of Copies Keep Stuff Safe) enables participating libraries to collect, store, preserve and provide access to their own local copies of content to which they have subscribed. The LOCKSS system was one of the two very different e-journal archiving approaches (the other being what is now known as Portico) supported by the Mellon Foundation in 2002. The LOCKSS application is open-source software. LOCKSS enables members to harvest the web-based presentation files of the content of e-journals to which they have subscribed from participating publishers. Access is triggered whenever (and for whatever reason) the material cannot be viewed on the publisher's (or intermediary's) servers. The highly distributed nature of the approach aims to ensure that there is sufficient replication to safeguard content despite any potential disasters which might be fall individual LOCKSS institutions. LOCKSS introduced the LOCKSS Alliance as a membership organization in 2005 to introduce governance to the program and address sustainability issues. In 2006, the JISC funded a two-year initiative, including a UK support post, to raise awareness of LOCKSS. A total of 24 institutions from around the UK joined the initiative, with a further 6 funding their own participation in the trial (Stanford University, LOCKSS Programme, 2013).

Pros

- It allows libraries to collect and exert control over the material they licence, as they have done when purchasing print journals.
- LOCKSS has increased publisher participation and the number of titles at a great rate.
- It covers a significant number of smaller and therefore probably more vulnerable publishers.
- It requires relatively modest investment in staff and equipment.
- It permits immediate access to the archive whenever there is a problem with communication with a publisher's server, even if very short term.

Cons

- It will need ongoing technical support.
- Although much technical support is managed remotely, staff from the subscribing institution needs to monitor the local server and ensure that it is running correctly.
- Future software development will require an active LOCKSS developer community.
- It may be difficult to integrate with other institutional technical platforms, e.g. several UK LOCKSS pilot reports referred to difficulties of proxy integration, though only

one appeared to think this would be a major barrier.

- Neither all titles of a particular publisher, nor all issues of a particular title are necessarily included in the LOCKSS collection.
- The larger STM publishers such as Elsevier have not demonstrated an enthusiasm for joining, and have expressed some concern about security of licensed content when it is distributed so widely. The resulting lack of major publisher content was a concern raised by several LOCKSS participants in their reports.

CLOCKSS

CLOCKSS "Controlled LOCKSS" was launched in 2006 and is still at a relatively early stage of development. A not-for-profit collaboration between libraries and publishers, it is a dark archive based on the LOCKSS software in which a limited number of libraries take on an archival role on behalf of a broader community. CLOCKSS ingests and saves either source or a presentation file as the publisher chooses. Each library hosts two servers which create a network of dark repositories. CLOCKSS intends to add a limited number of additional libraries and more publishers in the next phase. It describes itself as complementary to LOCKSS, and a "strategic component in a multilayer, resilient, local and international preservation plan." The 7 libraries participating in CLOCKSS are also members of the LOCKSS Alliance, 6 are U.S Based, and one is in the UK (University of Edinburgh).

Pros

- It provides a community approach, with a small number of distributed libraries assuming responsibility for long-term archiving, working in partnership with publishers.
- Low cost. The CLOCKSS website claims that "Fees are low and for a limited period while we build the CLOCKSS Endowment which we expect to underwrite costs after five years."
- It is likely to appeal to publishers with concerns about security of access and therefore attract a broader range of content from major publishers.

Cons

- Post-cancellation access is not supported.
- Long term costs are not clear, though there is an expressed commitment to reducing costs through the creation of an endowment.
- It is too early to be able to assess the long-term viability of the programme, including funding support.
- It is unclear what motivations (other than public good) CLOCKSS centres would have for continuing to participate in the programme long-term.

Portico

Portico is the second approach to e-journal archiving supported by the Mellon Foundation. It was launched as an independent organisation in 2005, though it has been in planning and preparation since 2002 under the auspices of Ithaka and with support from JSTOR.

Designed specifically as a third party service for scholarly literature published in electronic form, beginning with e-journals, it provides insurance to libraries that the e-journal content they have subscribed to will be preserved for the long-term. Portico provides access to the participating publisher is able to provide, Portico preserves normalised source files and/or presentation files of e-journals and also performs interpretation archiving. If no presentation files are provided, Portico creates new

presentation files. Portico can become a delivery mechanism in the event of a trigger event. In addition, if a publisher has designated Portico as such, it can also serve as a potential mechanism for post cancellation access.

Pros

- It removes the onus of managing the content from the library and provides assurance of long-term preservation.
- It can also provide post cancellation access providing the publisher has nominated them as a potential mechanism for this.
- Publisher participation has grown at an impressive rate. Some of the major STM publishers, such as Elsevier has joined.
- Their archiving approach is very thorough, preserving source files and/or presentation files, depending on what the publisher chooses to supply.
- The recent agreement with the KB to act as a mirror site for Portico strengthens their credibility.
- It provides an equitable business model, with publishers contributing to costs as well as libraries.

Cons

- Some see the dependence on publishers for revenue as a weakness.
- The fees libraries pay to Portico may be higher than other options.
- The current focus seems to be on the larger publishers Like CLOCKSS; it requires a cultural shift from libraries in terms of their traditional custodial role.
- Some believe that the title and publisher coverage could become rather US-centric once major international publishers have been covered.

e-Depot

The Koninklijke Bibliotheek (KB) is the national library of the Netherlands and operates e- Depot, its archive for the Dutch national deposit collection of electronic publications and other e-content (e.g. Dutch newspapers). e-Depot, which is OAIS compliant, was established in 2003 and focused initially on Dutch material. Recognising the international nature of journal publishing, this has now been extended to international publications. The KB intends to conclude archiving agreements for all the journals from 20-25 of the world's largest publishers. Publishers wishing to make use of the services provided by e-Depot are required to conclude an archiving agreement with the KB and to deliver bulk content and specified metadata. The primary deposit file format is PDF. Generally, end-user access is restricted to on-site perusal for reasons of private research only and on-line access is denied.

KB, who is committed to a programme of research and development in this area, is promoting the concept of the "Safe Places Network". This recognises the power and utility of international cooperation in finding solutions to the challenges of long-term preservation. A recent example is the decision by Portico to lodge a copy of their archive with KB.

Pros

- e-Depot aims to cover all major STM publishers. It already has more than 12 major publishers on board.
- The KB has a strong reputation for leadership in digital preservation research and practice.
- e-Depot saves both rendition and source files.
- There are no costs involved for others; the service is currently underwritten by the Dutch government. KB sales literature though includes the statement "In the coming years the KB intends to develop a

sustainable business model for the e-Depot which will reflect both the public and private responsibility for our digital and cultural heritage."

- They have demonstrated a good understanding of technical issues.

Cons

- Because they are such major publishers, the trigger events which could provide open access to content archived by e-Depot are highly unlikely to occur.
- It is extremely doubtful that the KB will want to engage in managing numerous authentication systems in the case of cancellation of content, the trigger event most likely to cause disruption for most organisations licensing access.
- At the moment, assured access is only available onsite at the KB.

Electronic Collections Online (ECO)

ECO was launched by OCLC in 1997 as a subscription service for libraries to a wide range of e-journals. It currently provides web access through OCLC's *First Search* service to over 5,000 titles from over 40 publishers. OCLC negotiates with publishers for perpetual access rights for subscribers to the service and for it to migrate back files to new formats if required.

Pros

- The emphasis is on continued access, which is likely to be of paramount concern to most libraries.
- It has been in operation longer than any of the others.
- It has a significant number of publishers and titles and good content coverage.

Cons

- Continued access is dependent on payment of an access fee. If titles are cancelled, access to past content can only be resumed if a subscription is reinstated within a five year period.
- ECO is primarily a tailored aggregator service for libraries: long term preservation is not their main mission.
- The emphasis on current access may be detrimental to long-term preservation (though some libraries find this feature makes the service easier to justify funding).

British Library e-journal Digital Archive

Over the past few years the BL has been building a Digital Object Management System (DOM) capable of storing and managing all digital content the BL takes responsibility for, including their own digitally created content, material purchased and material acquired through voluntary and legal deposit. The DOM system, which will be OAIS16 compliant, has three geographically dispersed identical nodes to provide redundancy in the event of loss of data from any one node.

The BL is currently assessing the feasibility of providing archiving and preservation solutions over and above those envisaged for e-legal deposit. The British Library began ingesting content from selected publishers during 2007, but is still finalising the exact service options that will be provided.

Access will depend on publishers' requirements. BL proposes two levels of service, Grey Archive (BL Reading Room access only), and Light Archive with content made available on the Internet worldwide following a publisher-specified trigger event (Maniatis, Petros and others, 2005).

Pros

- The British Library's strong reputation, experience and mandate for preservation generally;

- The archive is funded by charges to publishers and there are no other costs for libraries;
- The BL's extensive planning and preparation for digital deposit in building an infrastructure capable of dealing with both volume and complexity of material.

Cons

- It is too early to assess the content they will attract. As yet there is no detail of which publishers are on board. The emphasis is on recruiting publisher participation.
- The trigger events the BL use do not include post-cancellation access (apart from BL Reading Room access).

Criteria for judging relevance and value of new archiving initiatives

Factors which might impact on decision making for archiving solutions could include:

- Does the service include titles which are core to our institution's primary clientele/designated community?
- Does the overall bundle of titles in the archive meet our needs?
- Is our Library planning a move to e-only access?
- Does our Library intend to deselect print journals as electronic back files become available?
- Do we need assurance of continued *online* access?
- Do we need assurance of continued access for an indefinite period/longer than 10 years?
- Can we afford the service and does it represent value for money?

If the answer is yes to most of the above, then it will certainly be worthwhile investing in a trusted archiving solution, such as LOCKSS, Portico, and/or CLOCKSS.

Assuming the title is covered in more than one archive, what factors might affect which solution to go for?

- Do we want control over the titles that we have subscribed to in a similar way to print?
- Do we need seamless access following a trigger event?
- Do we want to delegate responsibility for archiving and perpetual access to a trusted third party?(e.g. Portico)
- Are we more concerned about the vulnerability of titles from smaller publishing houses rather than the major players?

If there is a combination of the content being of key importance and concern about the early stage of archiving solutions, it could be advisable to subscribe to more than one; assuming the content of primary interest is offered.

Selecting the right preservation strategy

There are various theories on the best way to preserve digital material, and a number of different approaches have been developed, which in turn have variants. They range from preserving the original technology on which the archival digital objects ran, to preserving only the significant properties of an object, which are defined independently of any specific hardware or software platform. Each approach has advantages and disadvantages.

Thibodeau suggests that a digital archive should take the following four criteria into consideration when selecting a preservation strategy:

- ***Feasibility:*** possession of hardware and software capable of implementing the chosen method.
- ***Sustainability:*** the method should be capable of being applied indefinitely into the future; or there should be another path which will offer a sequel to the method if it ceases being sustainable.
- ***Practicality:*** implementation should be within reasonable limits of difficulty and expense.
- ***Appropriateness***: the chosen approach should be appropriate for the particular types of digital objects to be preserved and the objectives of their preservation.

Some preservation methods only apply to specific hardware or software platforms, some to individual data types or formats, while others are very general. Depending on the range and variety of digital objects to be preserved by a repository, selection of approach might be limited to methods that are optimal for this range, or (if very wide ranging) a method with broad applicability should be chosen.

CONCLUSION

The problem of long-term digital preservation is becoming more real in the midst of a digital era. Old assumptions regarding information preservation are no longer valid, and it is clear that aggressive actions are needed to ensure the understandability of data for ages to come. In order to address these challenges, new technologies and systems are being developed. Such systems will be able to better address these vital issues if they are equipped with storage technology that is inherently dedicated to preservation and that supports the different aspects of the preservation environment. An appropriate storage system will make any solution more robust and lower the probability of data corruption or loss.
Although there is a trend towards “e-only” options, restraining factors include some academics distrust of e-publications and resistance to change. Drivers for a move to e-only include access, space, user expectations, and round the clock availability. It would be helpful if a cost-benefit analysis could be carried out comparing the savings from hardcopy cancellations (space, binding etc) with e-only licensing plus archiving costs. Which solution, if any, is chosen will depend on several factors including critical mass of content, numbers of publishers, numbers of titles, issue coverage, and costs. It should be acknowledged that publishers themselves may be able and willing to provide ongoing post-cancellation access, though this may be a chargeable service.

With the Digital Millennium Act in force, the publishers are not obligated to provide perpetual online access to content to which libraries used to have subscription. Implementing the LOCKSS/CLOCKSS concept in library can facilitate the perpetual access to such content. This, of course, can be done only through negotiation with the concerned publisher and it is likely to have financial implications.

Despite these agreements and developments, the field of digital archiving is still in its infancy, and much work needs to be accomplished to achieve a secure and permanent archiving of electronic journals.

Our study has examined different approaches to archiving and preservation. All deserve support and encouragement, but none currently offers the typical academic library a complete solution to their archival needs. Nor do any of them currently cover the greater proportion of the journal titles being published today.

REFERENCES

1. Beagrie, N. and Jones, M. (2002). "Trusted Digital Repositories: Attributes and Responsibilities", Mountain View, CA, Research Libraries Group- OCLC, available at http://www.rlg.org/longterm/repositories.pdf
2. "CLOCKSS", http://www.clockss.org **(**Accessed on 18 October 2014).
3. "Digital Millennium Act", available at http://www.copyright.gov.in.legislation/dmca.pdf (Accessed on 18 October 2014).
4. Digital Preservation Coalition (2002). Preservation Management of Digital Materials: A Handbook. Digital Preservation Coalition, New York. Available at www.dpconline.org/graphics/handbook/ (Accessed on 22 October 2014).
5. Hockx-Yu, Helen (2006). "Digital preservation in the context of institutional repositories", Electronic library and information systems. Vol. 40, No. 3, pp. 232-243.
6. Kirchhoff, Amy J. (2008). "Digital preservation: challenges and implementation", Learned Publishing, 21, pp. 285–294.
7. Lavoie, B. (2004). "The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making", White Paper, OCLC Research, available at: www.oclc.org/research/projects/digipres/incentives-dp.pdf (Accessed on 22 October 2014).
8. "LOCKSS", Wikipedia, available at http://en.wikipedia.org/wiki/LOCKSS **(**Accessed on 18 October 2014).
9. Maniatis, Petros [et al]. (2005), "The LOCKSS Peer-to-Peer Digital Preservation System", ACM Transactions on Computer Systems, 23, Issue 1, pp. 2–50.
10. McDonald, J. and Van de Velde, E. (2004). "The lure of linking", Library Journal, available at: www.libraryjournal.com/article/CA405398.html (Accessed on 08 October 2014).
11. Obaiah, B. and Jayakanth, Francis (2008). "Long-term preservation and perpetual access to scholarly content: the LOCKSS way", Proceedings of National conference on digitization and digital preservation, pp. 190-194.
12. Pool, Rebecca (2013). "Mind the gap: catching digital content before it slips away - PRESERVATION- Research Information" available at www.researchinformation.info/features/feature.php?feature_id=397 (Accessed on 16 October 2014).
13. Reich, V. and Rosenthal, D. (2001). "LOCKSS: a permanent Web publishing and access system**."** D-LIB Magazine, Vol. 7, No. 6. Available at: www.dlib.org/dlib/june01/reich/06reich.html (Accessed on 25 October 2014).
14. Rosenthal, David S. H.; Lipkis, Thomas Robertson; Thomas S. and Morabito, Seth (2005). "Transparent Format Migration of Preserved Web Content". D-Lib Magazine (Corporation for National Research Initiatives) 11 (1). doi:10.1045/january2005-rosenthal. (Accessed on 25 October 2014).
15. Stanford University LOCKSS Program (2013). "LOCKSS: Lots of Copies Keep Stuff Safe, Stanford University, Stanford". available at: www.lockss.org (Accessed on 15 October 2014).
16. http://www.paradigm.ac.uk/workbook/preservation-strategies/ (Accessed on 25 October 2014).
17. http://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf (Accessed on 25 October 2014).
18. https://www.jisccollections.ac.uk/Documents/e_journal_archiving_%20solutions_comparative_study.pdf (Accessed on 25 October 2014)

19. http://www.clir.org/PUBS/reports/pub107/pub107.pdf (Accessed on 25 October 2014).

Copyright © 2016, Satish Kumar. This is an open access refereed article distributed under the creative common attribution license which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.